USGS
science for a changing world

Data Management Plan framework (DMPf)

Community for Data Integration
January 12, 2011

Steve Tessler[1], Stan Smith[2], and Mike McHale[1]
[1]Climate Effects Network, [2]Alaska Science Center

U.S. Department of the Interior
U.S. Geological Survey

Data Management

USGS

Why manage research data?

- "Good data management is the foundation for good research. If data are properly organized, preserved and well documented, and their accuracy, validity and integrity is controlled at all times, the result is high quality data, efficient research, outputs based on solid evidence, and the saving of time and resources."

  – UK Data Archive

USGS

What is Data Management?

- "Data management is the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets." – DAMA (Data Administration &Management Association, Intl.) DMBOK (Data Management Body of Knowledge)
- Areas of Responsibility
  - Data governance
  - Data architecture, analysis, design
  - Database management
  - Data security
  - Data quality
  - Data integration
  - Data warehousing, business intelligence
  - Records management
  - Metadata management
  - Contact data management, personal information (maybe)

USGS

The DMPf Approach

USGS

DMPf Origins

- Framework was a collaboration between …
  - Mike McHale – Climate Effects Network
  - Steve Tessler – New Jersey Water Science Center
  - Stan Smith – Alaska Science Center
- We each needed to create a DMP for our respective Programs and wanted to share workload and facilitate data integration.
- Influenced by:
  - Gathering of existing DMPs and related documents from many federal agencies and research organizations world-wide.
  - Serving as the external reviewers for NEON's DMP and technical data models.
  - Multiple interactions with NPS and their *Inventory and Monitoring Data Management Plan*.

USGS

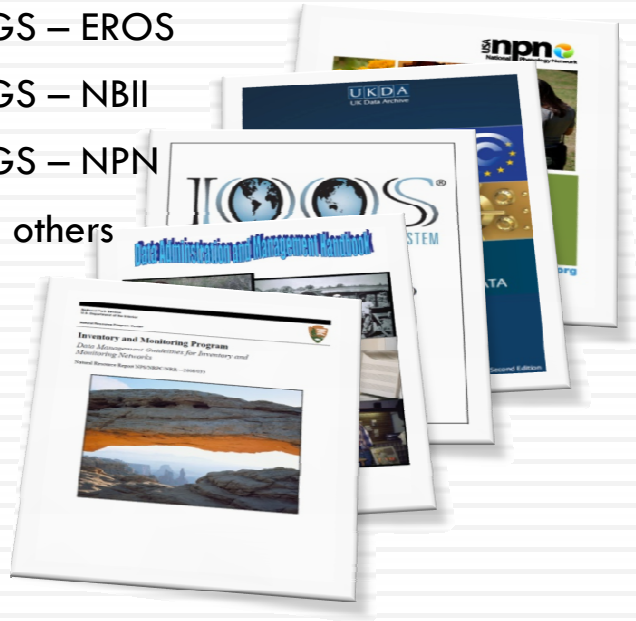
Objectives for the DMPf

- All the usual Data Management stuff … plus
  - Write DMP so that major sections of the work could be used by multiple programs (keep data management content isolated from program content)
  - Lay a foundation for true inter-program data integration
  - Bring the practice of managing data as an enterprise assets to USGS
  - Adapt the DMP to research organizations
  - Be comprehensive
  - Don't re-create – borrow from the best

USGS

Existing Data Management Plans

- Researched existing DM plans from …
  - LTER
  - NEON
  - Data Managers Association (DAMA)
  - University of Melbourne
  - Australian National Data Services (ANDS)
  - United Kingdom Data Archive
  - BLM
  - IOOS
  - NASA
  - NOAA
  - National Academy of Sciences
  - USGS – CERT
  - USGS – EROS
  - USGS – NBII
  - USGS – NPN
  - and others



USGS

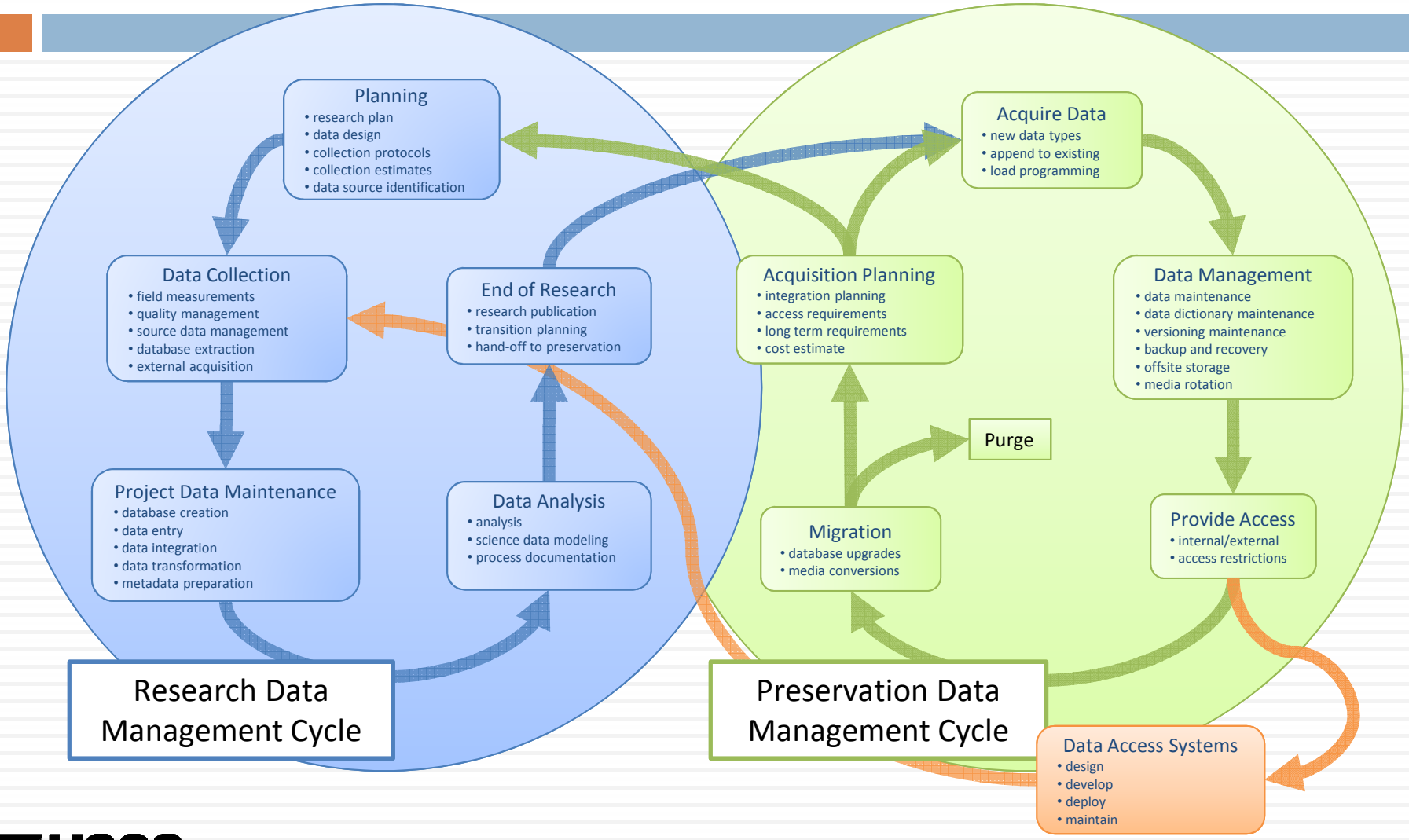
DMPf Concepts

- Research DM and Data Preservation
  - Proposed by UK Data Archive
- Data Levels
  - Used by NASA, NEON, and others
- Document Layering
  - Used by NPS Inventory & Monitoring

- What we gain from combining these concepts
  - A highly reusable framework that requires programs/projects only to write a slice of DMP.
  - Data integration achieved between projects and programs through shared standards inherited from foundation layers.
  - A clear and concise vocabulary for data management discussions.

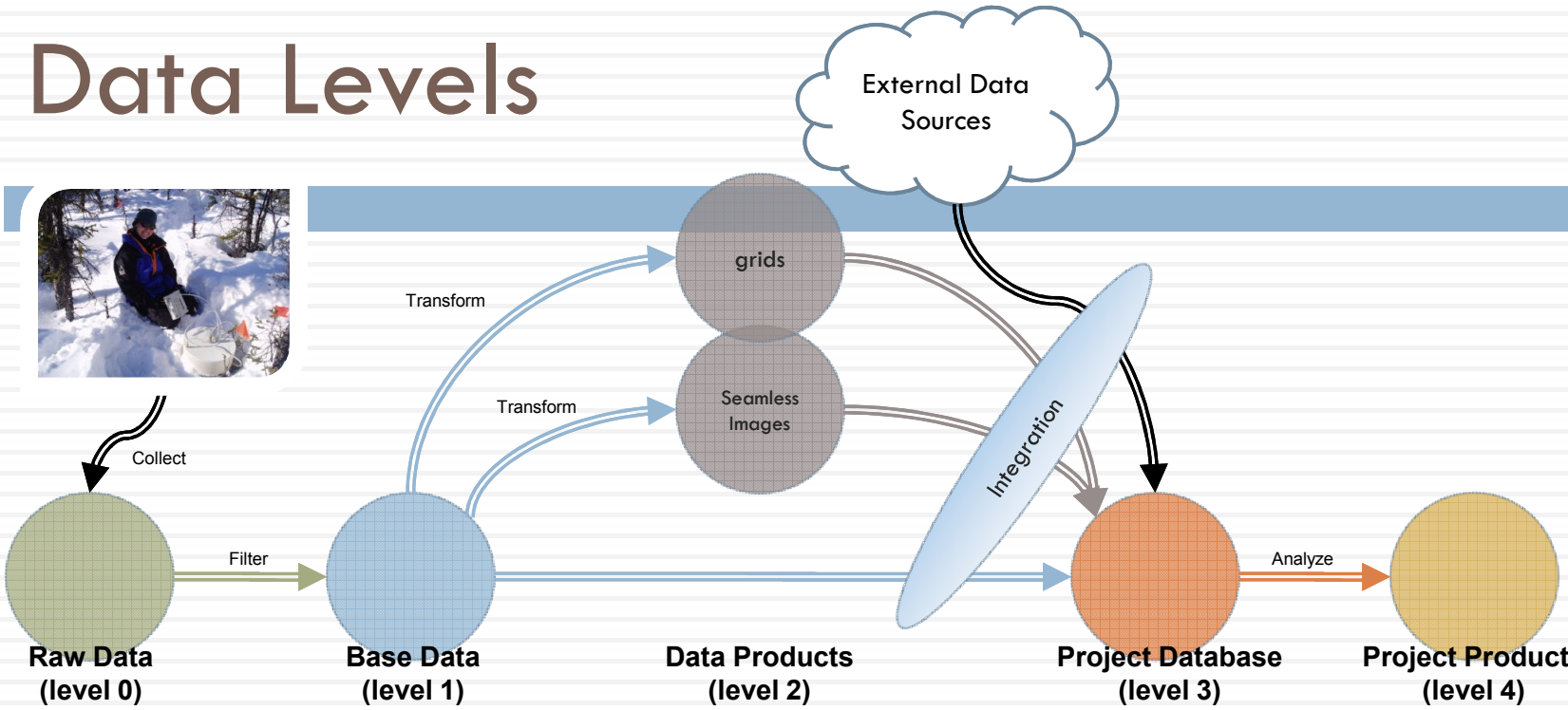
USGS

Research and Preservation



USGS

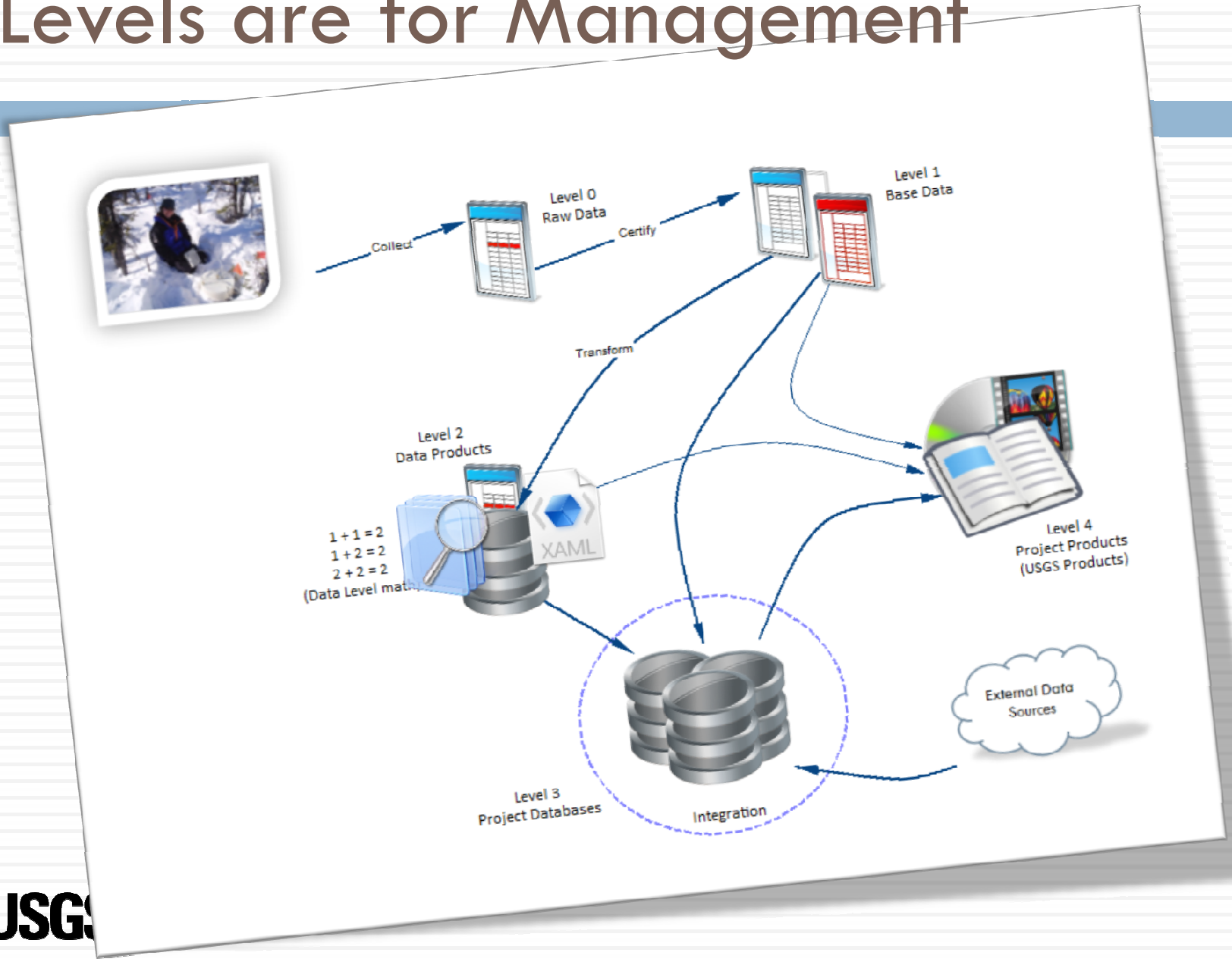
Data Levels



	Raw	Base	Products	Projects	Derived
Accessible	Archive	Online	Online	Project Team Only	Publication
Format	ASCII Text (CSV)	Relational DB	Various Technologies	Multi-format Proprietary software	Acrobat (.pdf) Print
Life Cycle	Archive media life	15+ year redesign	3-5 year tech life	Project life	Publication life
Metadata	Collection protocol Data quality Observer	Data model (ERD) Table/field def'n	Transform method Source tracing	Project Integration method Science methodology	Report Subject keyword

USGS

Levels are for Management



Document Layering

Data Management

- The things *common to all DMPs*
  - Concepts
  - Roles
  - Vocabulary
  - Glossary
  - …

Enterprise Requirements

- *What was decided*
  - Laws and Policies
  - Adopted standards
  - Standards adoption process
  - Intellectual property policy
- *The 'what' must be done*
  - QA/QC
  - Metadata
  - Backup
  - Alternate storage
  - …

Program Requirements

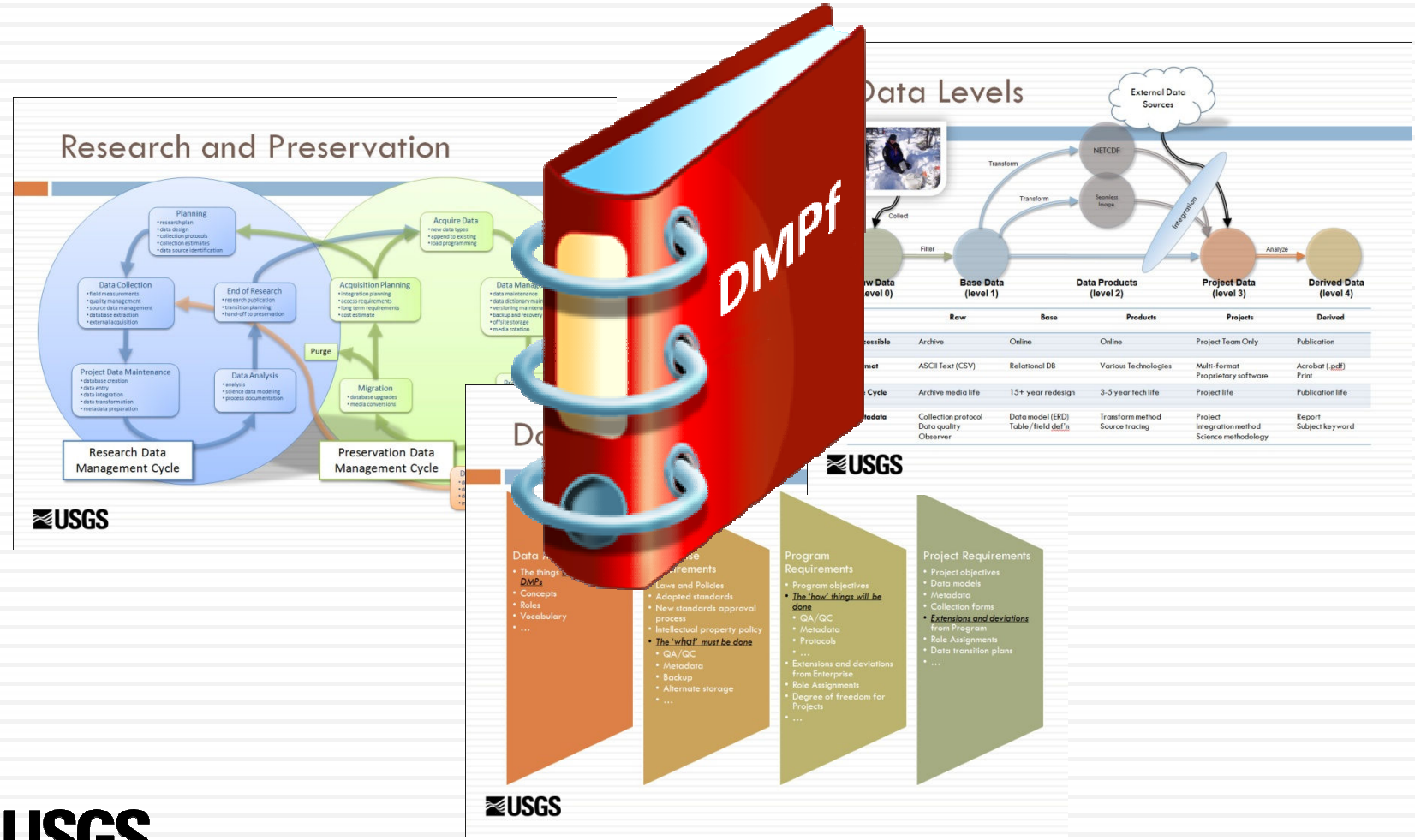
- Program objectives
- *The 'how' things will be done*
  - QA/QC
  - Metadata
  - Protocols
  - …
- Extensions and deviations from Enterprise
- Role Assignments
- Set degree of freedom for the Projects
- …

Project Requirements

- Project objectives
- *The 'details'*
  - Data models (ERD)
  - Metadata
  - Collection forms
  - Extensions and deviations from Program
  - Role Assignments
  - Data transition plans
  - …

USGS

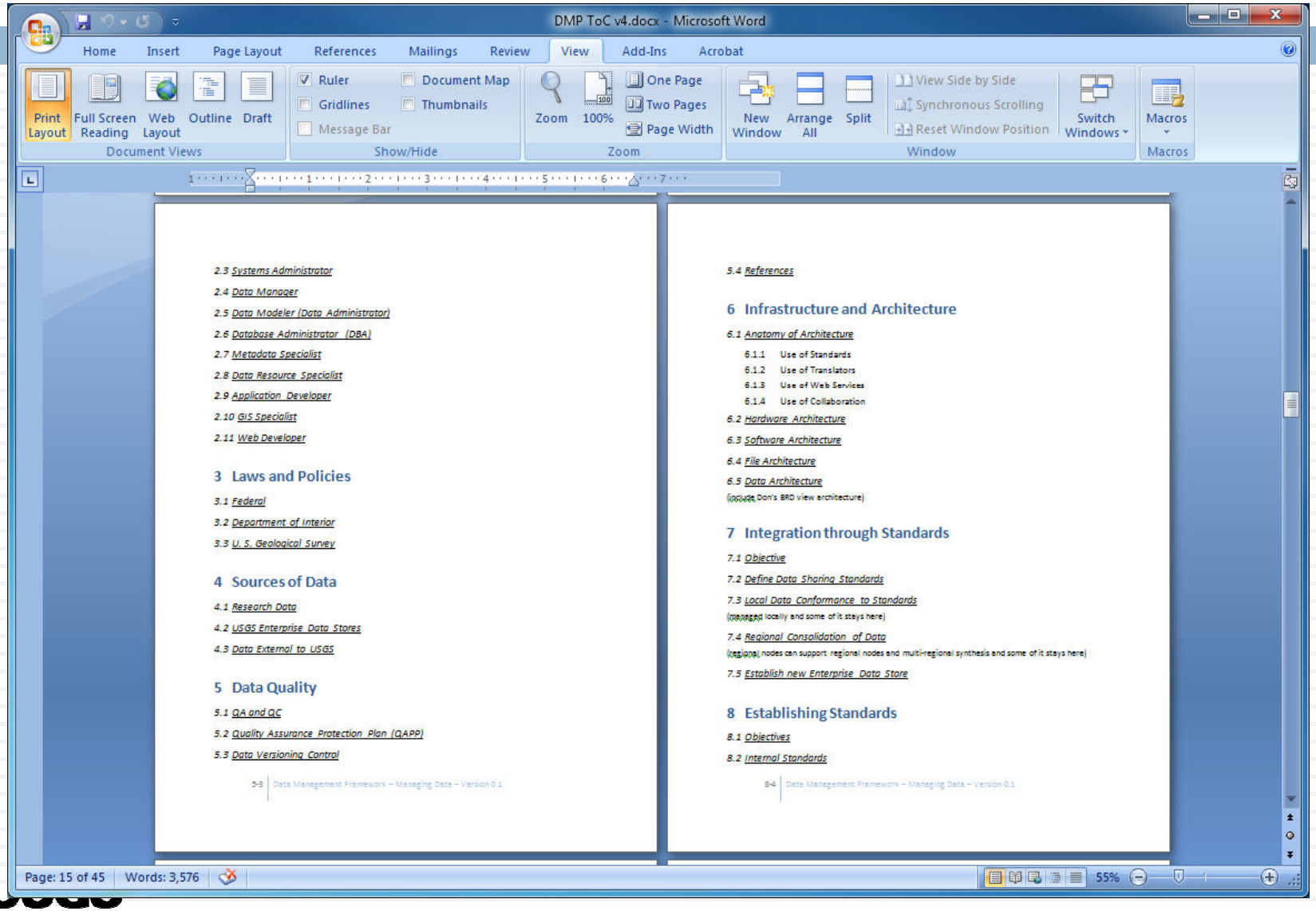
From Concepts to Framework



USGS

Framework Development Session

Framework Development Session



DMP Framework

I. Data Management
II. Standards
III. Research Data Management Plan (RDMP)
    a. Enterprise RDMP
    b. Program RDMP
    c. Project RDMP
IV. Preservation Data Management Plan (PDMP)
    a. Enterprise PDMP
    b. Program PDMP
V. Exposure and Delivery Plan (EDP)
    a. Enterprise EDP
    b. Program EDP
    c. Project EDP
VI. Appendix

Program Responsibility

For each topic in RDMP and PDMP framework

- Description of task
- Explanation of deviations or extensions to the Enterprise/Program RDMP
- List extension or deviations to Enterprise/Program standards
- Justify deviations
- Assign persons to roles
- Set degree of freedom for Programs/Projects

USGS

DMP Framework

I. Data Management
II. Standards
III. Research Data Management Plan (RDMP)
    a. Enterprise RDMP
    b. Program RDMP
    c. Project RDMP
IV. Preservation Data Management Plan (PDMP)
    a. Enterprise PDMP
    b. Program PDMP
V. Exposure and Delivery Plan (EDP)
    a. Enterprise EDP
    b. Program EDP
    c. Project EDP
VI. Appendix

Project Responsibility

For each topic in RDMP and PDMP framework

- Description of task
- Explanation of deviations or extensions to the Enterprise/Program RDMP
- List extension or deviations to Enterprise/Program standards
- Justify deviations
- Assign persons to roles
- Set degree of freedom for Programs/Projects

USGS

How to write the DMPf (1/2)

- Gradually
  - We do not need the full document completed to begin good data management practices
  - Add sections gradually as we work through multiple programs and projects
  - Assign work to most appropriate group
- Data Management - (DMPf Section I)
  - USGS Community for Data Integration (CDI) should accept the challenge to write this section for all of USGS – mostly lifted from existing DMPs and reference materials
- Standards - (DMPf Section II)
  - Build over time with contributions from DMP writing efforts of Programs and Projects USGS-wide
  - Core Science Systems (CSS) should manage the adoption, organization, and publication standards
  - Use PowerDesigner data dictionary repository for distribution of data design standards

USGS

How to write the DMPf (2/2)

- Enterprise Requirements - (DMPf Sections III.a, IV.a, V.a)
  - Initiate with CDI, Powell Center focus group
- Program Requirements - (DMPf Sections III.b, IV.b, V.b)
  - Written by data management personnel assigned to the Program
- Project Requirements - (DMPf Sections III.c, V.c)
  - Written by PI & science team with support from the Program data management personnel

USGS

Discussion

- A natural by-product of the DMPf approach is that it offers a true opportunity for USGS-wide integration of our data.

- Questions?

USGS